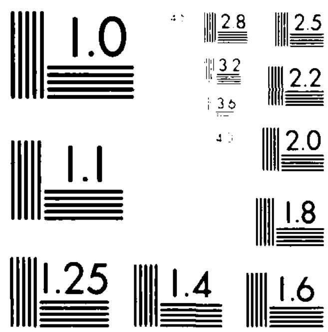
AD-A170 209 ON PREDICTION INTERVALS FOR FUTURE OBSERVATIONS FROM THE INVERSE GAUSSIAN (U) SOUTH CAROLINA UNIV COLUMBIA DEPT OF STATISTICS W J PADGETT ET AL JAN 86 TR-109 1/1

UNCLASSIFIED AFOSR-TR-86-0412 AFOSR-84-0156 F/G 12/1 NL



MICROCOPY RESOLUTION TEST CHART

(2)

AD-A170 209

DTIC ELECTE JUL 2 5 1986 S D

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: UNCLASSIFIED

1b. RESTRICTIVE MARKINGS:

2a. SECURITY CLASSIFICATION AUTHORITY: NA

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE: NA

3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for Public Release; Distribution Unlimited

4. PERFORMING ORGANIZATION REPORT NUMBER(S): Stat. Tech. Rep. No. 109 (62F25-1)

5. MONITORING ORGANIZATION REPORT NUMBER(S): AFOSR-TR- 86-0412

6a. NAME OF PERFORMING ORGANIZATION: University of South Carolina

6b. OFFICE SYMBOL *(If applicable)*:

7a. NAME OF MONITORING ORGANIZATION: AFOSR/NM

6c. ADDRESS *(City, State and ZIP Code)*:
Department of Statistics
Columbia, SC 29208

7b. ADDRESS *(City, State and ZIP Code)*:
Bldg. 410
Bolling AFB, DC 20332-6448

8a. NAME OF FUNDING/SPONSORING ORGANIZATION: AFOSR

8b. OFFICE SYMBOL *(If applicable)*: NM

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: AFOSR-84-0156

8c. ADDRESS *(City, State and ZIP Code)*:
Bldg. 410
Bolling AFB, DC

10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
6.1102F	2304	A5	

11. TITLE *(Include Security Classification)*: On Prediction Intervals for Future Observations from the Inverse Gaussian Distribution

12. PERSONAL AUTHOR(S): W. J. Padgett and S. H. Tsoi

13a. TYPE OF REPORT: Interim Technical

13b. TIME COVERED: FROM ______ TO ______

14. DATE OF REPORT *(Yr., Mo., Day)*: 1986 January

15. PAGE COUNT: 9

16. SUPPLEMENTARY NOTATION:

17. COSATI CODES

FIELD	GROUP	SUB. GR.

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*: Prediction; Monte Carlo simulations; First passage time distribution; Life testing; Reliability; Quality control.

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

The problem of predicting, on the basis of an observed sample of size n from an inverse Gaussian distribution, a future observation from the same distribution is discussed. Two prediction intervals that have been proposed in the literature, one of which is an approximate prediction interval, are compared using Monte Carlo simulations. The results indicate that in many of the simulated cases the approximate prediction interval is superior with respect to larger estimated coverage probabilities and smaller estimated mean lengths. This is true in particular for n at least 15 and for 95% and 99% prediction intervals.

DTIC FILE COPY

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐

21. ABSTRACT SECURITY CLASSIFICATION: UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL: Maj. Brian W. Woodruff

22b. TELEPHONE NUMBER *(Include Area Code)*: (202) 767-5027

22c. OFFICE SYMBOL: AFOSR/NM

AFOSR-TR- 86-0412

DEPARTMENT OF STATISTICS

The University of South Carolina
Columbia, South Carolina 29208

ON PREDICTION INTERVALS FOR FUTURE OBSERVATIONS FROM THE INVERSE GAUSSIAN DISTRIBUTION

by

W. J. Padgett* and S. H. Tsoi

University of South Carolina

Statistics Technical Report No. 109
62F25-1

January, 1986

Department of Statistics
University of South Carolina
Columbia, SC 29208

* Research supported by U.S. Air Force Office of Scientific Research under grant number AFOSR-84-0156 and U.S. Army Research Office under grant number MIPR ARO 139-85.

On Prediction Intervals for Future Observations

from the Inverse Gaussian Distribution

W. J. Padgett and S. H. Tsoi

Department of Statistics
University of South Carolina
Columbia, SC 2920

Key Words - Prediction; Monte Carlo simulations; First passage time distribution; Life testing; Reliability; Quality control.

SUMMARY

The problem of predicting, on the basis of an observed sample of size n from an inverse Gaussian distribution, a future observation from the same distribution is discussed. Two prediction intervals that have been proposed in the literature, one of which is an approximate prediction interval, are compared using Monte Carlo simulations. The results indicate that in many of the simulated cases the approximate prediction interval is superior with respect to larger estimated coverage probabilities and smaller estimated mean lengths. This is true in particular for n at least 15 and for 95% and 99% prediction intervals.

1. INTRODUCTION

The inverse Gaussian distribution has been proposed as a lifetime model, and its properties have been studied by Chhikara & Folks [1], Chhikara & Guttman [2], Padgett [5], and Padgett & Wei [7], among others. This model applies to accelerated life testing and repair time situations where early failures dominate, and it has a nonzero asymptotic failure rate [1].

Statistical prediction intervals have many applications in quality control and in reliability problems, and such intervals have been derived for the inverse Gaussian distribution by Padgett [6] and Chhikara & Guttman [2] independently. Padgett [6] proposed an approximate prediction interval for the mean of future observations from the inverse Gaussian distribution. Monte Carlo simulation results indicated that his approximate prediction interval performed very well with respect to coverage probabilities. Chhikara & Guttman [2] obtained exact prediction intervals for a single future observation from the inverse Gaussian from both a frequentist and a Bayesian viewpoint. Their frequentist approach did not always provide two-sided prediction intervals, however. In this note, Padgett's [6] approximate interval will be compared with Chhikara & Guttman's [2] exact frequentist approach based on Monte Carlo simulation results.

The pdf of the inverse Gaussian distribution appears in several forms [3]. The form used here is that given by Tweedie [9] with parameters μ and λ:

$$f(x;\mu,\lambda) = (\lambda/2\pi x^3)^{\frac{1}{2}} \exp[-\lambda(x-\mu)^2/2\mu^2 x],\ x > 0\ (\mu > 0, \lambda > 0).$$

The mean of this distribution is μ, and λ is a shape parameter. The variance is μ^3/λ, so μ is not a simple location parameter. It will be assumed in section 3 that both μ and λ are unknown.

Notation List

MLE	Maximum Likelihood Estimator
$\mu, \hat{\mu}$	mean of inverse Gaussian distribution and its MLE
$\lambda, \hat{\lambda}$	shape parameter of inverse Gaussian distribution and its MLE
$I(\mu,\lambda)$	refers to inverse Gaussian distribution with parameters μ,λ
γ	denotes probability level of prediction interval
n,m	size of current, future random samples
$X_1,\ldots,X_n$	current random sample from inverse Gaussian distribution
$Y_1,\ldots,Y_m$	future random sample from inverse Gaussian distribution; independent of $X_1,\ldots,X_n$
$\bar{X}_n$	$\sum_{i=1}^{n} X_i/n$, sample mean
$\chi^2_{\gamma}(\nu)$	γ Cdf point of chi-square distribution with ν degrees of freedom
$F_{\gamma}(\nu_1,\nu_2)$	γ Cdf point of *F-distribution with* (ν_1,ν_2) *degrees of* freedom

Other, standard notation is given in "Information for Readers & Authors" at the rear of each issue.

2. THE PREDICTION INTERVALS

The MLEs of μ and λ are $\hat{\mu} = \bar{X}_n$ and $\hat{\lambda}^{-1} = \sum_{i=1}^{n} (1/X_i - 1/\bar{X}_n)/n$ [1,9]. Tweedie [9] showed that $\bar{X}_n$ and $\hat{\lambda}$ are independent, $\bar{X}_n$ has $I(\mu,n\lambda)$ distribution, and $n\lambda/\hat{\lambda}$ has chi-square distribution with n-1 degrees of freedom. Since $\bar{Y}_m$ has $I(\mu,m\lambda)$ distribution, by a result of Shuster [8], $m\lambda(\bar{Y}_m-\mu)^2/\mu^2\bar{Y}_m$ has chi-square distribution with one degree of freedom. Thus, $(n-1)m\hat{\lambda}(\bar{Y}_m-\mu)^2/(n\mu^2\bar{Y}_m)$ has F-distribution with (1,n-1) degrees of freedom. For a value $0 < \gamma < 1$, then

$$P\left[\frac{(\bar{Y}_m-\mu)^2}{\mu^2\bar{Y}_m} \le \frac{nF_\gamma(1,n-1)}{m(n-1)\hat{\lambda}}\right] = \gamma. \tag{2.1}$$

If μ is <u>known</u>, (2.1) can be solved for an exact 100γ% prediction interval for $\bar{Y}_m$, the mean of m future observations from the inverse Gaussian distribution. However, in the more practical case that μ is unknown, Padgett [6] proposed an approximation to (2.1) which was simple and always produced a two-sided prediction interval for $\bar{Y}_m$ given by the roots of the quadratic equation

$$\bar{Y}_m^2 - c(X)\bar{Y}_m + \bar{X}_n^2 = 0, \tag{2.2}$$

$$c(X) = [(n+m)^2\,\bar{X}_n^2\,F_\gamma(1,n-1)/nm(n-1)\hat{\lambda}] + 2\bar{X}_n.$$

Denoting the smaller of the two roots by $L_1(X)$ and the larger by $U_1(X)$, an approximate 100γ% prediction interval for $\bar{Y}_m$ is $(L_1(X),U_1(X))$. For $m=1$, this method gives an approximate 100γ% prediction interval for a single "future" observation, Y_1, based on the "current" sample $X_1,\ldots,X_n$.

Chhikara & Guttman [2] obtained the exact 100γ% prediction interval for the single future observation Y_1 as $(L_2(X), U_2(X))$, where

$$L_2(X) = [V_1+V_2^{\frac{1}{2}}]^{-1},\ U_2(X) = [V_1-V_2^{\frac{1}{2}}]^{-1},$$

$$V_1 = 1/\bar{X}_n + nF_\gamma(1,n-1)/(2(n-1)\hat{\lambda}),$$

$$V_2 = (n+1)F_\gamma(1,n-1)/((n-1)\bar{X}_n\hat{\lambda}) + n^2F_\gamma^2(1,n-1)/(4(n-1)^2\hat{\lambda}^2).$$

Chhikara & Guttman point out that this procedure does not always provide two-sided intervals since there is a positive probability that the difference V_1-V_2 can be negative. In this case, only a lower one-sided interval is admissible.

Since the exact procedure of Chhikara & Guttman [2] might not yield a two-sided interval for Y_1, where Padgett's [6] approximation always gives one, it is of interest to compare the two approaches. This is done by Monte Carlo simulations in section 3.

3. COMPARISON OF THE PREDICTION INTERVALS

To compare Padgett's [6] approximate prediction interval for a single future observation, Y, from the inverse Gaussian with the prediction interval of Chhikara & Guttman [2], Monte Carlo simulations were performed to estimate the coverage probabilities and average widths of the intervals. The procedure for generating a random number from the inverse Gaussian distribution given by Michael, Schucany & Haas [4] was used. The simulations were performed as follows:

i. For given values of n, μ and λ, 1000 pairs of samples $(X_1,\dots,X_n)$, Y were generated.

ii. For each pair of samples, the $100\gamma\%$ prediction intervals for Y, $(L_i(X), U_i(X))$, $i=1,2$, were computed, and the lengths of the intervals and the number of intervals containing Y were obtained.

iii. The average interval lengths from the 1000 pairs of samples were computed and the proportions of intervals containing Y were obtained as estimates of the mean interval lengths and coverage probabilities, respectively.

iv. Steps i-iii were repeated for several values of n, μ, λ and for $\gamma = .90, .95, .99$.

In the simulations, when λ was small, a significant proportion (often as high as 90% for $\lambda = .25$) of the samples did not yield a two-sided interval from Chhikara & Guttman's [2] procedure. In addition, the estimated coverage probabilities for samples resulting in two-sided intervals were quite low for

small n. Tables 1-3 show some of the simulation results.

Surprisingly, as γ increases, the approximate prediction interval proposed by Padgett appears to be superior to the interval of Chhikara & Folks. In all of the cases simulated, Padgett's interval had larger estimated coverage probabilities and/or smaller estimated mean widths, and for larger n values had smaller estimated mean widths. Also, the estimated coverage probabilities for Padgett's interval were always close to γ.

Table 1

Simulation Results for $\gamma = .99$

			Average Width		Coverage Probability	
μ	γ	n	C&G	Padgett	C&G	Padgett
1	0.25	5	30.225	208.308	0.806	0.983
1	0.25	30	583.178	38.365	0.991	0.989
3	0.25	5	36.460	3430.014	0.933	0.977
3	0.25	30	1895.497	368.060	0.957	0.990
1	1	15	103.805	12.354	0.988	0.993
1	4	5	47.602	9.157	0.986	0.988
1	4	15	4.568	4.084	0.993	0.994
1	4	30	3.714	3.531	0.990	0.991
5	1	15	550.348	332.821	0.947	0.990
5	4	5	403.608	220.023	0.945	0.989
5	4	30	96.125	62.778	0.987	0.992

Table 2

Simulation Results for $\gamma = .95$

			Average Width		Coverage Probability	
μ	λ	n	C&G	Padgett	C&G	Padgett
1	0.25	5	25.999	73.393	0.828	0.942
1	0.25	30	849.066	21.850	0.945	0.962
3	0.25	5	28.096	1441.764	0.759	0.948
3	0.25	30	577.047	226.021	0.932	0.957
1	1	15	11.995	7.161	0.951	0.956
1	1	50	6.449	5.964	0.963	0.960
1	4	5	30.269	4.420	0.946	0.959
1	4	30	2.391	2.387	0.937	0.944
5	1	15	782.390	198.000	0.940	0.973
5	1	30	776.896	138.490	0.948	0.948
5	4	5	200.971	100.599	0.905	0.956
5	4	30	45.309	37.529	0.961	0.942

Table 3

Simulation Results for $\gamma = .90$

			Average Width		Coverage Probability	
μ	λ	n	C&G	Padgett	C&G	Padgett
1	0.25	5	30.249	46.122	0.705	0.897
3	0.25	5	24.420	739.880	0.738	0.895
1	4	15	2.089	2.027	0.903	0.893
5	4	15	51.448	31.853	0.903	0.887

REFERENCES

[1] R. S. Chhikara, J. L. Folks, "The inverse Gaussian distribution as a lifetime model," Technometrics, vol 19, 1977 Nov, pp 461-468.

[2] R. S. Chhikara, I. Guttman, "Prediction limits for the inverse Gaussian distribution," Technometrics, vol 24, 1982 Nov, pp 319-324.

[3] N. Johnson, S. Kotz, Distributions in Statistics. Continuous Univariate Distributions - I. Houghton-Mifflin Co., Boston, 1970.

[4] J. Michael, W. Schucany, R. Haas, "Generating random variates using transformations with multiple roots," American Statistician, vol 30, 1976 May, pp 88-90.

[5] W. J. Padgett, "Confidence bounds on reliability for the inverse Gaussian model," IEEE Transactions on Reliability, vol R-28, 1979 Jun, pp 165-168.

[6] W. J. Padgett, "An approximate prediction interval for the mean of future observations from the inverse Gaussian distribution," J. Statist. Comput. Simul., vol 14, no 3, 1982, pp 191-199.

[7] W. J. Padgett, L. J. Wei, "Estimation for the three-parameter inverse Gaussian distribution," Commun. Statist. - Theor. Meth., vol A8, 1979 Aug, pp 129-137.

[8] J. Shuster, "On the inverse Gaussian distribution function," J. Amer. Statist. Assn., vol 63, 1968 Dec, pp 1514-1516.

[9] M.C.K. Tweedie, "Statistical properties of inverse Gaussian distributions. I," Ann. Math. Statist., vol 28, 1957 Jun, pp 362-377.

END

DTIC

8 — 86